BCI End

5th module

Brain signals → AC to DC →

Set of features → feature vector

Three steps of feature extraction:-

(i) Signal conditioning to reduce noise
(ii) Extraction of features
(iii) feature conditioning to properly prepare the feature vector for the feature translation stage

First step :- Signal condition

→ removing noise

Signal conditioning can include a number of diff procedures:-

(i) frequency range prefiltering
(ii) data decimation & normalization
(iii) Spatial filtering

Second step :- Block processing

→ For BCI applications, it is highly desirable for the processing

Time (Temporal) features → Peak Pick
Correlation & Template matching

Peak Picking

→ determines minimum and maximum value of the signal samples in a specific block of time and uses that value as feature for the time block.

→ Signal can be averaged or integrated over all or part of the time block to yield feature of the block.

Extracting the features
2) Correlation and template matching

Frequency features

band power, FFT, autoregressive modelling

Translational algorithms → feature conditions

Normalization, log normal transforms, feature smoothing

PCA & ICA; removing irrelevant and redundant features
(Maximum relevant Minimum redundant)

Feature translation

Third step : feature conditioning

translation algorithms

---

feature translation

Discriminant functions, regression functions

Regressional analysis

The regressional analysis is a statistical method to deal with formulatio of mathematical model depicting relationship among variables which can be used for purpose of predicton.

Classification of regression analysis models

Linear regression models.

- simple linear regression
- multiple linear regression

Least square method

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

We are to minimize the value of SSE and hence to determine parameters of a & b

$$\frac{\partial(SSE)}{\partial a} = -2\sum_{i=1}^{n}(y_i - a - bx_i)$$

$$\frac{\partial(SSE)}{\partial b} = -2\sum_{i=1}^{n}(y_i - a - bx_i)\,x_i$$

For minimum value of SSE

$$\frac{\partial(SSE)}{\partial a} = 0 \qquad \frac{\partial(SSE)}{\partial b} = 0$$

$$\boxed{a = \bar{y} - b\bar{x}} \qquad \boxed{b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

R^2: measure of quality of fit → coefficient of determination

SST → total corrected sum of squares

$$= \sum_{i=1}^{n}(y_i - \bar{y})^2$$

$R^2 = 1 - \frac{SSE}{SST}$; $R^2 = 1$ means good fit

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Baysean classifier (Assumptions : mutually exclusive & exhaustive)

statistical classifier

Bayseian classifier is an approach for modelling probabilistic relationships between the attribute set & the class variable.

M estimate

$$P(A_j = a_j / C_i) = \frac{n_{c_i} + mp}{n + m}$$

n = total number of instances from class C_i

n_{c_i} = number of training examples from class C_i that takes $A_j = a_j$

m = it is a parameter known as the equivalent sample size,

p = it is a user specified parameter.

~~Vertebrate classification~~

Entropy

To deal with classification job, entropy ~~the~~ is an important concept.

an information - theoretic measure of uncertainty contained in training data.

$$E = -\sum_{i=1}^{m} p_i \log_2 p_i$$

$$-\frac{4}{24} \log \frac{4}{24} - \frac{5}{24} \log_2 \frac{5}{24} - \frac{15}{24} \log_2 \frac{15}{24}$$

Decision tree induction is top-down, recursive, divide & conquer approach.

ID3

At each node, the splitting attribute is selected to be the most informative among the attributes not yet considered in path starting from the root.

Weighted entropy

$E_A(D) \Rightarrow$ expected information ~~gain~~ required to classify a tuple from D based on splitting of A.

$$E_A(D) = \sum_{j=1}^{m} \frac{|D_j|}{|D|} E(D_j)$$

$\frac{|D_j|}{|D|}$ denotes the weight of the j^{th} training set.

Info gain

$$\alpha(A,D) = E(D) - I_A(D)$$

$E(D)$ ↓ Total dataset

$I_A(D)$ ↓ For A = 1, A = 2, ... A = n' class

The ID3 strategy of attribute selection is to choose to split on the attribute that gives greatest reduction in the weighted average entropy.

α frequency table

Info gain ≥ 0

→ When the entropy of training set takes largest value ($\log_2 k$) (this occurs when the classes are balanced) then the info gain will be zero.

CART Algorithm

CART is a technique that generates a binary decision tree

Gini index of diversity known as γ

Weighted average gini index $= G_A(D) = \frac{|D_1|}{D} G(D_1) + \frac{|D_2|}{D} G(D_2)$

$V(A,D) = G(D) - G_A(D)$

$1 - \sum p_i^2$

CART → for more classes

Algorithm C4.5

On limitation of ID3

If attribute has distinct values for all tuples, then it would result in large number of partitions, each one containing just one tuple.

ID3 may suffer from overfitting problem.

$$\text{Gain ratio} = \frac{\text{Information gain}}{\text{split information}}$$

Linear regression

Least squares estimation

$$\min \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{m} (y_i - (w \cdot x + b))^2$$

$$\hat{w} = (X^T X)^{-1} X^T Y$$

$$\text{LASSO} \rightarrow \min \sum_{i=1}^{m} (y_i - w \cdot x_i - b)^2 + C \sum_{j=1}^{n} |w_j|$$

regularization term

$$\text{Ridge regression:- } \min (y_i - w \cdot x_i - b)^2 + C \sum_{j=1}^{n} |w_j^2|$$

Soft margin

$(x_i, y_i) \quad i = 1 \ldots m$

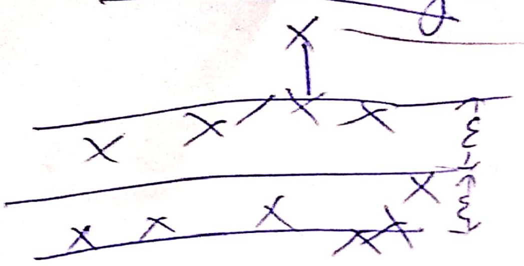
minimize

$$\frac{1}{2}\|w\|^2 + C \sum_{i=1}^{m} (\xi_i + \xi_i^*)$$

Under constraints

$$\begin{cases} y_i - (w \cdot x_i) - b \leq \varepsilon + \xi_i \\ (w \cdot x_i) + b - y_i \leq \varepsilon + \xi^* \\ \xi_i , \xi_i^* \geq 0 , i = 1 \ldots m \end{cases}$$

Support Vector regression



$$L_\varepsilon(y, f(x,w)) = \cancel{\max}$$
$$= \max\left(|y - f(x,w)| - \varepsilon, 0\right)$$

Given training data

(x_i, y_i) ; $i = 1 \dots m$

minimize

$$\frac{1}{2}\|w\|^2 + C \sum_{i=1}^{m} (\xi_i + \xi_i^*)$$

Neural Networks

$$\delta_K \leftarrow y^K - w \cdot x^K$$

$$w_i \leftarrow w_i + \alpha\, \delta_K\, x_i^K$$

Perceptron Node

$$\Delta w_i = (t - z) \times c \times x_i$$

wt from input i to perceptron node, c is the learning rate, t is the target for the current instance, z is the current output, and x_i is i^{th} input.

Perceptron Convergence theorem: Guaranteed to find a solution in finite time if a solution ~~set~~ exists

$$w_{new} = w_{old} + \overbrace{\alpha}^{\text{learning rate}} \underbrace{(y_i - \bar{y}_i)}_{\text{error}}$$

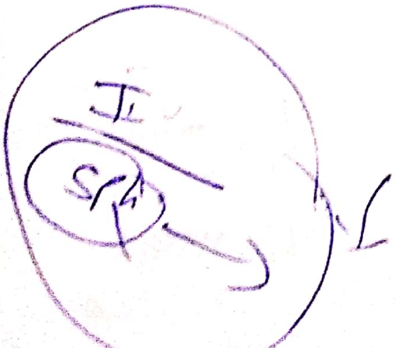
weights will be updated according to errors

10 µm

12×10^{-5} m

1 mm $\quad 12 \times 10^{-2}$ mm

$$\frac{0.12 \text{ mm}}{1 \text{ mm}}$$



3rd module

Invasive

→ microelectrode → electrode made of tungsten, platinum or iridium alloy

→ Intracellular recordings → measures voltage or current across the membrane of the neuron.

→ Tetrodes and multi-unit recording.

→ The most common type of implantable arrays are micro-, silicon-based, and flexible microelectrode arrays.

→ To record larger number of neurons, microelectrodes can be arranged in a grid like structure to form a multielectrode array of m × n electrodes

Partially invasive

ECoG

↓

ECoG electrodes can record the electrical fluctuations caused by coherent activity of large populations of neurons.

Micro ECoG

Optical recording : Voltage sensitive dyes

→ Two photon calcium imaging

Based on the fact that electrical activity in neurons with ~~flourescent calcium - indicator of dyes~~. changes in Ca concentration.

Photon Ca imaging involves :-

(i) using pressure ejection to load neurons with flourescent Ca - indicator dyes.

(ii) Monitoring changes in Ca flourescence during neural activity using two photon microscopy.

Non Invasive

EEG signals reflect the summation of thousands of post synaptic potentials

Temporal resolution is good, spatial is poor

α → 8 to 13 ; β → 13 to 30

γ → 30 to 100 ; δ waves → 0.5 - 4 ; θ 4-8

α - 8 - 13 , β - 13 to 30 γ - 30 to 100

δ - 0.5 to 4 θ → 4 to 8

MEG → magneto encephalography (better spatial resolution than EEG)

fMRI → Bold → paramagnetic properties

fNIR → measures changes in blood oxygenation level caused by increased neural activity in brain.
Based on detecting near-infrared light absorbance of haemoglobin in the blood with & without oxygen

Positron Emission Tomography

measures emissions from radioactively labeled, metabolically active chemicals that have been injected into the blood stream for transportation to the brain.

The labelled compound is called radio tracer.

SPECT single photon emission computed Tomography

↓
uses gamma rays

Principals to discriminate artifacts from EEG signals

Physiological activity has a logical topographical field of distribution with an expected fall of voltage potentials.
Artifact have an illogical distribution that defies the principal of localization.

Cognitive Workload

CWL is a demand placed upon humans for mental resources while performing a task.
mental resources include working memory, ability to process etc

Working memory :- is a cognitive system with a limited capacity to hold a small amount of information & process it.